**Information Retrieval**

**10 February 2020**

**Ex 1 [ranks 4+4]** Assume that you are given a set of strings S = {bat, bit, bet, but}.

* Build the data structure that efficiently searches for an arbitrary pattern P with 1 edit distance.
* Show how it is executed the 1-edit search for P = “best”

**Ex 2 [ranks 4+4]** Assume you are given the following Elias-Fano encoding of a sequence of integers:

L = 11 10 00 01 11 00 01 11

H = 0 0 1 0 1 0 1 1 1 0 1 1 0 0 1 0

* Show and explain which is the **number of integers** encoded, and which is the **number of bits** used by the original encoding for each integer
* Decompress the 4th integer

**Ex 3 [points 4+4]** You are given the files: F\_old = “cane gatto orso”, F\_new = “matto cane gas”, and assume a block size B=3 chars.

* Show the execution of the algorithm rsync. *(comment the various steps)*
* Show the execution of the algorithm zsync. *(comment the various steps)*

**Ex 4 [points 3+3]** Answer the following questions

* Comment on the difference in building and querying between the Term-based versus Document-based partitioning in Distributed indexing.
* Given two nodes in a directed graph, comment on how we can measure their “closeness” by taking into account the graph structure, and explain the difference with computing the shortest path.